Hetergoneity in Lifetime Earnings Risk

Ethan Ballou*

June 10, 2025

Abstract

Abstract. This is our abstract. It is abstract.

Keywords:
JEL Codes:

*University of Wisconsin - Milwaukee

1 Introduction

This is an example citation [**?**].

Multiply all Coefficients by 100?

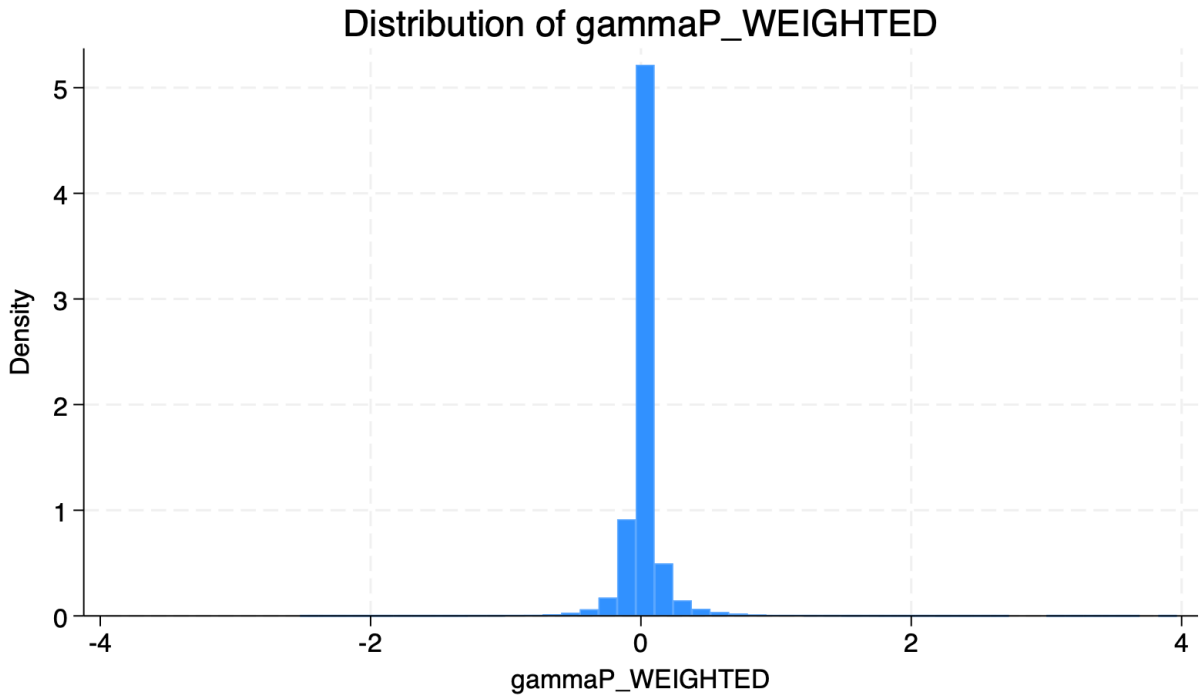

Figure 1: Distribution of gammaP_WEIGHTED

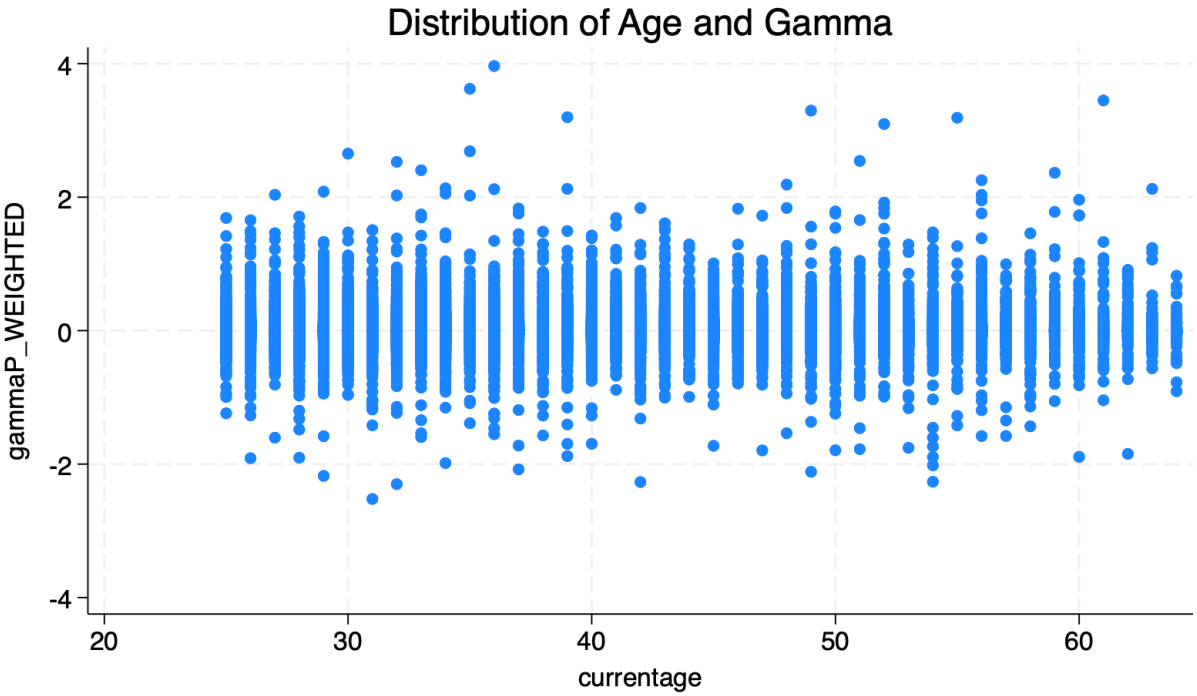

Figure 2: Scatterplot of Age vs. gammaP_WEIGHTED

Table 1: OLS Estimates for γ (Coefficients $\times$ 100)

	(1)	(2)	(3)	(4)	(5)
EDU1	-0.361	-0.360	-0.522^{*}	-0.597^{*}	-0.630^{**}
	(0.255)	(0.273)	(0.286)	(0.307)	(0.310)
EDU2	-0.283	-0.315^{*}	-0.418^{**}	-0.417^{*}	-0.455^{**}
	(0.175)	(0.181)	(0.193)	(0.216)	(0.219)
EDU3	-0.089	-0.063	-0.138	-0.114	-0.147
	(0.206)	(0.209)	(0.216)	(0.228)	(0.230)
PrRecess	-0.006	-0.038	-0.037	-0.040	-0.039
	(0.005)	(0.036)	(0.036)	(0.036)	(0.036)
rGDPgrow	-0.033	0.076	0.071	0.063	0.058
	(0.033)	(0.172)	(0.172)	(0.172)	(0.172)
fhwage0_P0	-0.006	0.004	0.005	0.002	0.004
	(0.025)	(0.027)	(0.027)	(0.028)	(0.028)
ma5aep	0.004	0.003	0.004	0.004	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.004)
veteran	0.021	-0.003	0.053	0.018	0.040
	(0.141)	(0.151)	(0.153)	(0.154)	(0.155)
OLF	0.627	0.584	0.550	0.619	0.625
	(0.627)	(0.628)	(0.628)	(0.629)	(0.629)
tenure	-0.010	-0.009	-0.008	-0.012	-0.010
	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
currentage	0.874^{**}	0.936^{**}	0.926^{**}	0.919^{**}	0.907^{**}
	(0.380)	(0.384)	(0.385)	(0.385)	(0.385)
currentagesq	-0.023^{**}	-0.025^{***}	-0.025^{***}	-0.025^{***}	-0.024^{***}
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
currentagecube	0.0002^{***}	0.0002^{***}	0.0002^{***}	0.0002^{***}	0.0002^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Occupation Controls				✓	✓
Industry Controls			✓		✓
Other Controls		✓	✓	✓	✓

Notes: Standard errors in parentheses. Other controls include state, year, race, and cohort fixed effects. Statistical significance: $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$. All coefficients and standard errors are multiplied by 100 for easier interpretation.

Table 1 shows the OLS estimates of gamma using the different specifications.

- The coeffiecnts are quite small. However the gamma is centered very tigthly around 0 - The only coefficents are that are signifcant are EDU1, EDU2, currentage, currentagesq, and currentagecube. - EDU1 and EDU2 are only somewhat significant, EDU2 being more significant with insutry controls - Education as a trend at least is negative in all cases. - Age is positive and signiicant while the squared term is negative and more significant.

Table 2: Stepwise Results for γ

	(1)	(2)	(3)	(4)	(5)
EDU1	selected	2	2	selected	selected
EDU2	selected	1	1	selected	selected
EDU3	6	10	8	6	6
PrRecess	1	3	3	1	1
rGDPgrow	4	4	4	4	4
fhwage0_P0	7	11	13	9	9
ma5aep	2	6	6	3	3
veteran	8	12	12	10	11
OLF	3	5	5	2	2
tenure	5	7	9	5	5
currentage	selected	selected	selected	selected	selected
currentagesq	selected	selected	selected	selected	selected
currentagecube	selected	selected	selected	selected	selected
Occupation Controls	-	-	-	selected	selected
Industry Controls	-	-	7	-	10
Cohort Controls	-	8	10	7	7
Race Controls	-	13	14	11	12
Year Controls	-	9	11	8	8
State Controls	-	selected	selected	selected	selected
Occupation Controls				✓	✓
Industry Controls			✓		✓
Other Controls		✓	✓	✓	✓

Notes: This table reports results from stepwise regression models using a p-value threshold of 0.05. "Selected" indicates variables retained in the final model. Numbers indicate the order of variable removal (with 1 being the last variable removed before model finalization). "-" indicates the variable was not included in the initial model specification.

Table 2 shows the stepwise estimates of gamma using the different specifications.

- Similar to OLS, All the age variables have explnatory power, and EDU1 and EDU2 are mildy explanatory in some cases. - Occupation and state controls are selected in all cases - Probability of recession, and real GDP growth, are top 4 every time - veteran, wages, and many of the controls are seen as the least important variables in just about all the models

Table 3: Lasso Results for γ

	(1)	(2)	(3)	(4)	(5)
EDU1	3	3	2	1	1
EDU2	2	1	1	1	1
EDU3	8	7	6	6	4
PrRecess	4	Not Selected	Not Selected	Not Selected	Not Selected
rGDPgrow	6	Not Selected	Not Selected	Not Selected	Not Selected
fhwage0_P0	7	9	9	8	6
ma5aep	1	2	1	2	1
veteran	10	8	8	7	5
OLF	3	2	3	1	1
tenure	5	4	4	3	2
currentage	9	6	7	5	4
currentagesq	11	10	10	8	7
currentagecube	2	5	5	4	3
Occupation Controls	-	-	-	Selected	Selected
Industry Controls	-	-	Selected	-	Selected
Cohort Controls	-	Selected	Selected	Selected	Selected
Race Controls	-	Selected	Selected	Selected	Selected
Year Controls	-	Selected	Selected	Selected	Selected
State Controls	-	Selected	Selected	Selected	Selected
Occupation Controls				✓	✓
Industry Controls			✓		✓
Other Controls		✓	✓	✓	✓

Notes: This table reports variables selected by Lasso regression with Bayesian Information Criterion (BIC) variable selection. "Selected" indicates variables retained in the final model. Numbers in parentheses indicate the order in which variables were added to the model. "-" indicates the variable was not included. "Not Selected" indicates the variable was not selected by Lasso but was provided in the model specification.

Table 3 shows the lasso of gamma using the different specifications.

- The are quite different from the stepwise results. All controls are selected in all cases - PrRecess and rGDPgrow are not selected in any case across lambda values except in the first model - Non of the continuous variables are selected in any of the models once the optimal lambda is selected by cv - Howvever the EDU1 and EDU2 are the most important varaibles after the selected varaibles in most cases - MA5 and OLF are also quite strong and come after EDU1 and EDU2 in most cases - In the lasso models the age varaibles aren't selected and in some cases are the last variables considered across lambda values

Table 4: Lasso and SHAP Results for Occupations

Occupation	SHAP Rank	LASSO Rank	Occupation	SHAP Rank	LASSO Rank
occ_21	1	Not Selected	occ_60	2	16
occ_84	3	26	occ_61	4	13
occ_1	5	14	occ_45	6	20
occ_70	7	Not Selected	occ_98	8	14
occ_37	9	Not Selected	occ_97	10	27
occ_2	11	20	occ_99	12	22
occ_13	13	14	occ_9	14	24
occ_11	15	23	occ_83	16	21
occ_4	17	3	occ_101	18	Not Selected
occ_95	19	25	occ_79	20	15
occ_85	21	Not Selected	occ_6	22	8
occ_999	23	Not Selected	occ_8	24	Not Selected
occ_93	25	17	occ_55	26	21
occ_20	27	16	occ_53	28	23
occ_17	29	6	occ_19	30	Not Selected
occ_5	31	18	occ_40	32	23
occ_58	33	26	occ_34	34	19
occ_59	35	13	occ_7	36	Not Selected
occ_77	37	21	occ_87	38	21
occ_50	39	19	occ_14	40	Not Selected
occ_54	41	28	occ_96	42	15
occ_38	43	20	occ_3	44	12
occ_15	45	14	occ_32	46	14
occ_18	47	11	occ_42	48	11
occ_73	49	Not Selected	occ_31	50	Not Selected
occ_62	51	21	occ_44	52	17
occ_33	53	19	occ_63	54	28
occ_49	55	18	occ_86	56	17
occ_36	57	Not Selected	occ_43	58	16
occ_39	59	Not Selected	occ_74	60	17
occ_72	61	Not Selected	occ_64	62	2
occ_56	63	Not Selected	occ_92	64	Not Selected
occ_80	65	Not Selected	occ_88	66	13
occ_12	67	14	occ_75	68	25
occ_81	69	Not Selected	occ_30	70	Not Selected
occ_35	71	9	occ_57	72	Not Selected
occ_89	73	Not Selected	occ_71	74	Not Selected
occ_16	75	29	occ_94	76	25
occ_82	77	Not Selected	occ_52	78	Not Selected

Notes: This table reports occupations selected by Lasso regression with Bayesian Information Criterion (BIC) for predicting earnings risk. "SHAP Rank" shows the variable importance ranking based on SHAP values (lower numbers indicate greater importance). "LASSO Order" indicates the order in which variables would enter the model if the penalty were relaxed. Note that the BIC-optimal model contained no occupation variables.

Table 5: Lasso and SHAP Results for Industries

Industry	SHAP Rank	LASSO Rank
twoind_3	1	28
twoind_9	2	10
twoind_19	3	18
twoind_30	4	2
twoind_16	5	19
twoind_21	6	8
twoind_14	7	Not Selected
twoind_18	8	Not Selected
twoind_5	9	7
twoind_33	10	3
twoind_10	11	21
twoind_4	12	27
twoind_29	13	21
twoind_999	14	17
twoind_15	15	16
twoind_7	16	13
twoind_11	17	16
twoind_12	18	3
twoind_22	19	27
twoind_1	20	11
twoind_25	21	10
twoind_27	22	3
twoind_6	23	23
twoind_20	24	11
twoind_23	25	18
twoind_8	26	28
twoind_24	27	29
twoind_31	28	23
twoind_28	29	7
twoind_13	30	25

Notes: This table reports industries selected by Lasso regression with Bayesian Information Criterion (BIC) for predicting earnings risk. "LASSO Selection Order" indicates the order in which variables would enter the model if the penalty were relaxed. "SHAP Ranking" shows the variable importance ranking based on SHAP values (lower numbers indicate greater importance). Note that the BIC-optimal model contained no industry variables.

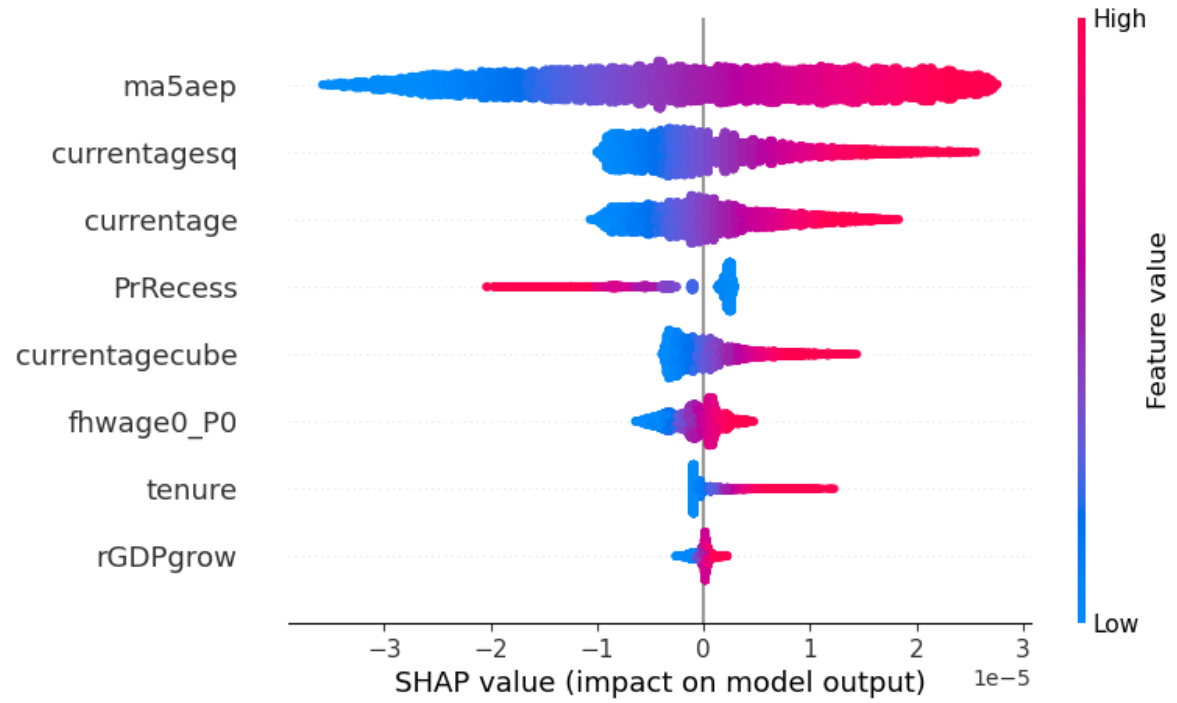

Figure 3: SHAP Summary Plot

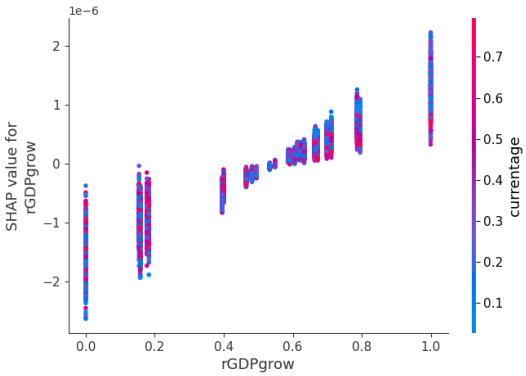

Figure 4: GDP by Age

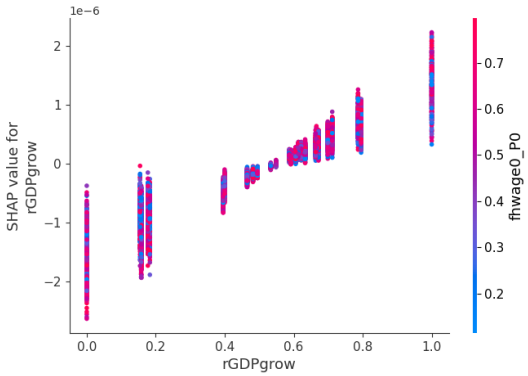

Figure 5: GDP by Income

References